

quinn emanuel trial lawyers | san francisco

50 California Street, 22nd Floor, San Francisco, California 94111-4788 | TEL (415) 875-6600 FAX (415) 875-6700

September 19, 2024

VIA ECF

The Honorable Jon S. Tigar
United States District Court for the Northern
District of California
Oakland Courthouse, Courtroom 6 - 2nd Floor
1301 Clay Street
Oakland, CA 94612

Re: **Letter Brief Regarding NVIDIA's ESI Protocol and Protective Order Proposals
Pursuant to ECF No. 68**

*Nazemian, et al., v. NVIDIA Corporation, Case No. 4:24-cv-01454-JST
Dubus, et al., v. NVIDIA Corporation, Case No. 4:24-cv-02655-JST*

quinn emanuel urquhart & sullivan, llp

ABU DHABI | ATLANTA | AUSTIN | BEIJING | BERLIN | BOSTON | BRUSSELS | CHICAGO | DALLAS | DOHA | HAMBURG | HONG KONG | HOUSTON | LONDON | LOS ANGELES | MANNHEIM | MIAMI | MUNICH | NEUILLY-LA DEFENSE | NEW YORK | PARIS | PERTH | RIYADH | SALT LAKE CITY | SAN FRANCISCO | SEATTLE | SHANGHAI | SILICON VALLEY | SINGAPORE | STUTTGART | SYDNEY | TOKYO | WASHINGTON, DC | WILMINGTON | ZURICH

Dear Judge Tigar:

Defendant NVIDIA Corporation (“NVIDIA”) in the above-captioned cases respectfully requests that the Court adopt NVIDIA’s proposed protective order and ESI protocol.

I. The Court Should Adopt NVIDIA’s Proposed Protective Order

NVIDIA respectfully requests entry of the “Model Protective Order for Litigation Involving Patents, Highly Sensitive Confidential Information and/or Trade Secrets” (the “Model Order”) with minor revisions tailored to the needs of this case. This case concerns a family of NVIDIA large language models (“LLMs”), collectively referred to as the NeMo Megatron series. Plaintiffs are seeking disclosure of proprietary, sensitive, and highly-technical information, such as source code and associated data used to implement and train the NeMo Megatron series. To accommodate this discovery, NVIDIA proposes the Model Order with minor revisions to reflect the issues here and to address NVIDIA’s source code production environment.

Plaintiffs’ proposal departs from the Model Order by (1) rewriting the definitions of Attorneys’ Eyes Only and Source Code to drastically reduce the materials that qualify for protection, (2) deeming each current NVIDIA employee the supposed “author” of any document created by any other current or former NVIDIA employee, (3) allowing any expert the right to review NVIDIA’s Source Code without advance disclosure to NVIDIA and the opportunity to object, and (4) providing an artificial and unworkable process for addressing the clawback of inadvertently produced privileged materials. Given these unreasonable changes to the Model Order, NVIDIA respectfully requests that that Court enter NVIDIA’s proposed protective order.

A. NVIDIA’s Proposed Protective Order Tracks the Model Order

NVIDIA proposes only three edits to the Model Order, each designed to address the facts of this case. First, NVIDIA proposes adding “non-public machine learning and/or large language

models” and “non-public or proprietary model training data” to the definition of Source Code in Section 2.9 of the Model Order, leaving the rest of the definition unchanged. An LLM is stored as a data structure that includes the logic, billions of parameters that are uniquely generated through training, and information that describes the logical interconnection of the individual processing nodes within the model. If one obtained improper access to a non-public model, with some knowledge of its architecture and parameters, one could use it as a foundation on which to build their own, shortcircuiting the extensive development efforts required for the original. Accordingly, both the training data and the models reflect sensitive algorithms and software architectures analogous to source code.

Second, NVIDIA’s proposal adds another option for secure source code review. To protect against improper transmission of source code, Section 9(c) of the Model Order prohibits the source code review computer from having a connection to the Internet. NVIDIA’s source code review process accomplishes the same objective in a different way; the source code review computer has a connection to the Internet but one that can be used only to securely access a source code review repository hosted at NVIDIA. Adding this option for hosting source code will have no impact on Plaintiffs and will help NVIDIA ensure the efficient and secure review of its source code.

NVIDIA’s proposal also includes some restrictions on devices that may be used when reviewing source code to protect against unauthorized copying and a process for ensuring the availability of resources for source code review. These measures reasonably ensure the protection of NVIDIA’s highly sensitive and proprietary code. Plaintiffs raised no objections to these procedures in prior meet-and-confer correspondence, though their current proposal no longer contains NVIDIA’s proposed changes.

Third, NVIDIA’s proposal adds one sentence to Section 13 of the Model Order to note that the parties agree that inadvertent production of privileged material does not constitute a waiver of any otherwise applicable privilege. Plaintiffs do not appear to disagree with this statement, and it is consistent with Federal Rule of Evidence 502(d).

B. The Court Should Reject Plaintiffs’ Proposed Protective Order, Which Undermines the Model Order’s Protections

Plaintiffs propose four areas of edits that unnecessarily weaken the Model Order’s protections and add needless, complex procedures directed at hypothetical disputes.

1. Plaintiffs’ Proposal Materially Narrows the Model Order’s Source Code and AEO Protections (Sections 2.8, 2.9, and 2.18)

Plaintiffs’ proposal narrows the definitions of “Attorney’s Eyes Only” and “Source Code” with no justification. Plaintiffs limit “Attorney’s Eyes Only” protection to only two categories: confidential research and development; and financial, technical, or other sensitive trade secret information. The Model Order, however, protects any “extremely sensitive ‘Confidential Information or Items, disclosure of which to another Party or Non-Party would create a substantial risk of serious harm that could not be avoided by less restrictive means,” without limitation to any arbitrary categories. Model Order § 2.8. Plaintiffs’ proposed definition may unduly exclude personal information of NVIDIA employees, non-public contracts or agreements with third parties, or other sensitive categories.

Plaintiffs’ definition of “Source Code” is also unworkably narrow, limiting the information qualifying for heightened protections to “human readable set of instructions written in a programming language.” This proposal removes categories of information that the Model Order recognizes warrant the same protection as “computer code,” such as “associated comments and revision histories, formulas, engineering specifications, or schematics” that can “describe or

otherwise describe in detail the algorithms or structure of software or hardware designs.” It also appears to exclude confidential and proprietary LLMs and training data, which will be at the core of this case. The Model Order acknowledges the importance of protecting Source Code by covering both the written computer language that makes up “source code” and associated information that could reveal proprietary and highly sensitive competitive information. Courts consistently recognize the extreme sensitivity of Source Code and grant it careful protection. *Drone Techs., Inc. v. Parrot S.A.*, 838 F.3d 1283, 1300, n.13 (Fed. Cir. 2016) (“it is well recognized among lower courts that source code requires additional protections to prevent improper disclosure because it is often a company’s most sensitive and most valuable property”); *Via Vadis Controlling GmbH v. Skype, Inc.*, No. 12-MC-193, 2013 WL 646236, at *3 (D. Del. Feb. 21, 2013) (noting “[s]ource codes are the most sensitive and confidential property of Respondents. When disclosed in U.S. litigation, extreme measures are ordered to protect their confidentiality”).

Plaintiffs’ proposal fails to recognize that other documentation, like architectural or design documentation, may reveal details of how source code operates that is tantamount to revealing the source code itself. As the Model Order recognizes, parties need a mechanism to protect such sensitive information. The theoretical possibility of over-designation (Plaintiffs’ sole justification for their narrow definition) is already addressed by the Model Order’s robust procedure for challenging improperly designated materials. Model Order § 6.

2. Plaintiffs’ Proposal Deems All NVIDIA Employees “Authors” of Any Document Produced by NVIDIA for Evidentiary Purposes, Unrelated to the Protective Order (Sections 7.2(g), 7.3(f) & (g))

Plaintiffs have proposed language that would deem every NVIDIA employee to be the “author” of any document created by *any other current or former* NVIDIA employee. Plaintiffs’

expansive language would remove requirements for evidentiary authentication and foundation, and seeks to impart all of NVIDIA’s knowledge to any individual employee regardless of whether the employee has seen the document in question. Plaintiffs’ attempt to circumvent the rules of evidence has no place in this case, let alone in the protective order.

3. Plaintiffs’ Proposal Removes NVIDIA’s Ability to Review Expert Backgrounds and Object Prior to Disclosure of Its Confidential Information (Section 7.4(a)(2))

Plaintiffs propose to remove the Model Order’s protections against sharing confidential information with third-party experts until the experts have been disclosed and subject to objections (if appropriate). The Model Order requires parties to identify a proposed expert and their past business and professional associations and contemplated future work over a five-year period and provides for objections and relief if there are potential conflicts. This procedure protects against disclosure of information to individuals that could use that information outside the litigation adversely to the producing party, now or in the future. Plaintiffs’ proposal, in contrast, does not even require notice to NVIDIA before sharing NVIDIA’s Source Code or Attorney’s Eyes Only information. Instead, the proposal would prevent only access by a current or anticipated “officer, director, or employee” of a “competitor,” as determined by Plaintiffs, contrary to the case law. *GPNE Corp. v. Apple Inc.*, No. 5:12-CV-2885-LHK-PSG, 2014 WL 1027948, at *2 (N.D. Cal. Mar. 13, 2014) (disqualifying an expert under protective order where expert was an “active consultant in the field at issue,” and “the information to which he would be exposed . . . could influence his work as a consultant . . . ”). Plaintiffs’ sole justification for this proposal is that it would make it “easier” to share information with their experts. That does not justify depriving NVIDIA of the opportunity to intervene where necessary to protect its confidential information.

4. Plaintiffs’ Proposal Adds an Onerous Procedure for Clawbacks (Section 16 Clawback & Rule 502(D) Order)

Plaintiffs' clawback proposal would require the Parties to undertake an onerous and unnecessary process in the event of inadvertently produced privileged materials that is not found in the Model Order and not warranted here. Federal Rule of Civil Procedure 26(b)(5)(B) already provides a procedure addressing the inadvertent production of privileged documents. Plaintiffs do not explain why a different process is needed. Requiring the parties to complete document review and witness preparation thirty days before a deposition, so that any need to clawback documents would be discovered more than thirty days before a deposition, is not feasible. Plaintiffs' proposal would delay discovery, including the unnecessary cancellation and rescheduling of depositions.

II. The Court Should Adopt NVIDIA's Proposed ESI Protocol

Consistent with Section G of the Court's Standing Order, NVIDIA proposes using the "Model Stipulation & Order Re: Discovery of Electronically Stored Information for Patent Litigation" with limited revisions in this case ("Model ESI Order"). The Model ESI Order is designed to provide for fair, efficient, and proportional discovery in complex, technical cases.

As set out above, this case turns on the technical details of the "training" of NVIDIA's NeMo Megatron models. The Complaint asserts one claim, for direct copyright infringement, based on NVIDIA's alleged use of certain data in training those models, and NVIDIA argues, *inter alia*, that any such use of Plaintiffs' works is a fair use under 17 U.S.C. § 107. Given the allegations, discovery will be focused on the technical details of NVIDIA's collection and transformation of data, its subsequent use in training of NVIDIA's NeMo Megatron models, and whether such use is fair. Many key facts are likely to be undisputed.

Given the technical nature of the case, NVIDIA's proposal is based on the Model ESI Order. But NVIDIA has also adopted Plaintiffs' proposals where possible.¹ NVIDIA's proposal further includes a small number of modifications shown in redline in the Declaration of Rachael L. McCracken. In contrast, as detailed below, Plaintiffs dramatically alter the Model ESI Order to increase the scope of discovery far out of proportion to the needs of this case.

A. A Model Order With Limited Modifications Is Practicable In This Case

The Model Orders are based on the experience of judges and parties conducting discovery in this District. The Court's Standing Order directs the parties to use the Model Orders where practicable (Standing Order § G) and NVIDIA used the Model Order for Patent Cases. Other courts, including this Court, have used the Model Order for Patent Cases in similar cases. *See, e.g., Doe v. Github, Inc.*, Case No. 4:22-cv-06823-JST (Nov. 1, 2023) Dkt. 168. While Plaintiffs claim to have used the Standard Model Order as a starting point, their extensive and burdensome revisions render it unrecognizable and should be rejected. Courts in this District (including this Court) have rejected broad proposals like Plaintiffs' in favor of proposals that more closely track the Model Orders. *Doe v. GitHub, Inc. et al*, Case No. 4:22-cv-06823-JST, Dkt. 165, Dkt. 168 (where parties' offered competing proposals, adopting ESI Proposal provided by Defendant that closely tracked N.D. Cal. Model Order and made limited modifications); *In re OpenAI ChatGPT Litigation*, 3:23-cv-03223-AMO; Dkts. 166, 175 (rejecting in part Plaintiffs' requested number of ESI custodians and proposal to allow the requesting party to have input in determining search

¹ NVIDIA has accepted Plaintiffs' changes to the following sections with limited modifications: Section 12 – Hit Reports ; Section 15 – Required Metadata Fields; Section 16 – Embedded Files; Section 17 – Hyperlinks; Section 18 – Privilege Logs; Section 19 – Model Data; and Section 20 – Documents Protected From Discovery.

terms).² Based on the public docket, no disputes have yet been raised to the Court based on the scope of the *Github* order.

B. Plaintiffs’ Proposed Number of Custodians and Search-Term Process (Plfs.’ Sections 5(A-C, G)) Are Extremely Burdensome

Plaintiffs’ proposed ESI protocol would require that each party identify 24 custodians for ESI discovery (nearly 5 times the number in the Model ESI Order) and reserves the right to add even more custodians. Plaintiffs’ Section 5(A). In this case, email is not expected to provide meaningful evidence on whether copyrights were infringed or whether NVIDIA’s alleged use of the data in training qualifies as fair use. Although NVIDIA’s proposed order compromises by providing for eight custodians, anything more than that would be disproportionate and wasteful.

Plaintiffs’ proposal also diverges substantially from the Model ESI Order in addressing search terms. The Model ESI Order, adopted nearly verbatim in NVIDIA’s proposal, provides for five search terms per custodian per party, articulates reasonable guidelines for constructing search terms and narrowing search criteria, and makes clear that outside the search-term process in the Order, “General ESI production requests under Federal Rules of Civil Procedure 34 and 45 shall not include email or other forms of electronic correspondence.” Model ESI Order ¶¶ 6, 11. Plaintiffs’ proposal, on the other hand, sets no limits on the number, form, or content of search terms and grants an unlimited ability to object to the Producing Party’s search terms. Plaintiffs’ Section 5(B)-(C). And it articulates no process for refining search terms, beyond a fourteen-day period to meet and confer, after which a party can “immediately submit the dispute to the Court or its designee for resolution.” *Id.* Plaintiff’s proposal also guts the search-term process by further

² Similarly, the parties in other cases involving similar technology have stipulated to orders containing provisions far closer to NVIDIA’s proposal. *UAB “Planner5D” v. Meta Platforms, Inc. et al*, Case No. 3:19-cv-03132-WHO, Dkt. 25; *Kadrey et al v. Meta Platforms, Inc.*, Case No. 3:23-cv-03417-VC, Dkt. 101.

requiring the parties to conduct an additional, undefined search of custodians' ESI that does not hit on any search terms. Plaintiffs' Section 5(G). Other courts (including this Court) have rejected such broad proposals. *DOE I et al v. GitHub, Inc. et al*, Case No. 4:22-cv-06823-JST, Dkt. 168 (selecting ESI Proposal provided by Defendant that closely tracked N.D. Cal. Model Order); *In re OpenAI ChatGPT Litigation*, 3:23-cv-03223-AMO; Dkts. 166, 175.

C. There Is No Basis to Require Custodian Mobile Devices and Social Media Data (Plfs.' Section 5(F))

Plaintiffs also propose invasive discovery of custodians' mobile phones, tablets, and computers, as well as their personal social media accounts. Unlike the Model ESI Order, which limits discovery to a custodian's email, Plaintiffs also seek discovery into voicemails, text messages and/or iMessages, private chat messages, calendars, photographs, videos, and/or social media accounts. Plaintiffs' Section 5(F). Plaintiffs' proposal is not proportional to the needs of this case, imposes irrelevant requirements, and creates undue burden by requiring the collection and assessment of irrelevant personal data that has no bearing on copyright infringement allegations. Plaintiffs' provision on social media data also contemplates subpoenaing NVIDIA's employees personally for production of "social media materials that are not within the producing Party's possession, custody or control." *Id.* at "Social Media Data." Plaintiffs provide no justification for this burdensome invasion of privacy.

D. Plaintiffs' Structured Data Process Is Impracticable and Unnecessary

In Section 11, Plaintiffs also propose a cumbersome process for the collection and production of "structured data." Plaintiffs' proposal is unnecessary, as the parties have already separately addressed procedures for source code and model-related data and Plaintiffs have not identified any other structured data relevant to this case. Plaintiffs have also done nothing to justify

their burdensome process, which would require extensive meet and confers where Plaintiffs could seek witness testimony from NVIDIA employees.

* * * * *

NVIDIA has addressed the major departures in Plaintiffs' proposal, but there are a number of additional unwarranted deviations from the Model Order that could not be addressed specifically or in detail given the extent of Plaintiffs' proposed edits and the applicable page limits.³

III. Conclusion

NVIDIA respectfully requests that the Court enter its proposed protective order and ESI protocol, which closely track the Model Orders.

Respectfully submitted,

QUINN EMANUEL URQUHART & SULLIVAN, LLP

By: /s/ Sean S. Pak
Sean S. Pak
Attorney for Defendant NVIDIA
Corporation

³ Plaintiffs make a number of additional, unwarranted additions to the Model Order that could not be addressed in detail given the breadth of their changes, including provisions that could be read to call for disclosure of attorney work product, such as the "methods to validate" the document review (*id.* Sec. 5(E)) and an explanation of the process of determining which ESI documents are produced (Sec. 5). These invasive terms should be rejected in their entirety.